深度学习在大豆叶片图像数据管理中的识别与分类研究

陆丽娜[1,2]，于　啸[1,2*]

（1. 山东理工大学 管理学院，淄博　255000；2. 山东理工大学 计算机科学与技术学院，淄博　255049）

摘　要：［目的 / 意义］为了提高大豆叶片图像的分类精度与效率，进一步对大豆叶片图像进行存储与管理。［方法 / 过程］本文利用深度学习方法，针对肉眼观察准确率较低且不同人群分类结果差异较大的大豆叶片图像数据提出了一种自动分类方法。本研究首先对大豆叶片进行 ROI 感兴趣区域划分，进而利用分水岭分割方法对大豆叶片进行提取，最后通过深度学习高效精确的实现了大豆叶片的分类识别。［结果 / 结论］通过分析大豆叶片形态图像特点后，基于深度学习开展了对大豆叶片形态的分类识别的研究，达到了较高的识别准确率。

关键词：深度学习；农业科学数据；数据分类；图象识别

中图分类号：G250.7　　文献标识码：A　　文章编号：1002-1248（2023）02-0087-08

引用本文：陆丽娜, 于啸. 深度学习在大豆叶片图像数据管理中的识别与分类研究[J]. 农业图书情报学报, 2023, 35(2)：87-94.

1 引　言

虽然大量文献表明中国农业数据的整理、发布和共享都还处于初级阶段[1]，但随着数据科学和计算机技术的深入发展，技术在数据管理上的应用更加广泛和深入。深度学习在文本数据识别和图像数据识别与分类方面都起到了重要作用，替代传统的机器学习方法，在大规模样本分析、识别与分类中，采用了无监督学习方法，本文结合大豆的图像数据识别与分类进行分析，通过案例分析，站在实验数据的图像数据采集视角，总结了大规模样本数据的农业科学数据采集与分类，从科学实验数据采集与分类的应用场景出发，展示了深度学习技术的应用进展。

在过去的 10 年中，图像处理和计算机视觉方法已广泛用于植物病害的检测和分类。在国内方面，学者们已经在植物表型数据处理方面发现了深度学习的成功应用，如翁杨等的基于深度学习的农业植物表型研究综述[2]，岑海燕等的深度学习在植物表型研究中的应用现状与展望[3]；同时，基于卷积神经网络的植物分类和品种识别也取得了一些研究成果，袁培森等基于卷积神经网络的菊花花型和品种识别[4]，吴剑涛等基于深

收稿日期：2021-03-15

基金项目：国家社会科学基金项目“大数据环境下农业科学数据监管机制构建研究”（18BTQ062）；山东省自然科学基金资助项目“植物表型数据管理的多学科交叉理论、方法与应用研究”（ZR2022MG047）

作者简介：陆丽娜（1977-），女，博士，副教授，研究方向为科学数据管理、管理科学与工程、信息系统与信息管理

*通信作者：于啸（1976-），男，博士，副教授，研究方向为农业信息化、数据库管理、数据科学。Email：neaufish@163.com

度学习的遥感图像苹果树冠信息提取，避免了实地测量，而且允许种植者动态监测果园树木的生长情况[5]。国外方面，已经建立了很多类似的植物病害的检测的数据集，比如柑橘类水果，叶子和茎的图像数据集[6]、PlantVillage 数据集、kaggle 数据集等。同时，还有很多基于深度学习植物图像的分类研究：MAYRA 等使用 3D-CNN 生成了整个研究区域的墙到墙树种地图，改进的树种分类可以有益于可持续林业和生物多样性保护[7]，MADS 等以杂草分类为例，利用深度卷积神经网络对植物物种分类[8]。

总之，深度学习在农业科学数据管理的应用主要集中在大规模数据的植物及植物表型的图像识别与分类，农业的病虫害检测、农作物及杂草检测与分类、农作物的产量预测等研究中[9]。粮食作物中的病虫害和各种疾病严重影响主产，并给全球经济造成重大损失。粮食作物健康的监测和疾病的早期诊断是可持续农业的关键任务，现有的谷物植物病害自动检测和分类方法仍处于起步阶段。因此，新颖的全自动工具对于谷物植物中疾病的检测和分类过程必不可少[10,11]。本文以大豆的图像数据的识别为例，探讨深度学习方法是如何在大豆叶片图像数据管理实践中进行应用的。

2 基于深度学习的农业科学图像数据识别与分类

2.1 数据准备

大豆叶片形态是大豆最重要的形态特征和生物学特征之一。正确地采集大豆叶片图像数据，可以提高深度学习分类识别的准确率，提高大豆叶片形态识别模型的泛化能力。

2.1.1 大豆叶片数据特点

大豆叶片形态以植株三出复叶中的顶小叶形状为准，分为披针、卵圆、椭圆和圆 4 种形态，如图 1 所示。数据主要具有以下特点。

（1）不同形态差异小：大豆叶片形态的分类与特征差异较为明显的多种植物叶片之间的分类不同，不同大豆叶片形态之间差异较小，人为区分难度较高。

（2）同种形态特征变化大：同一种类型的大豆叶片之间，形态特征变化较大，不具有典型性。

（3）叶片弯曲不平整：大豆叶片大多蜷缩不平整，容易影响人们对叶片形态的判断。

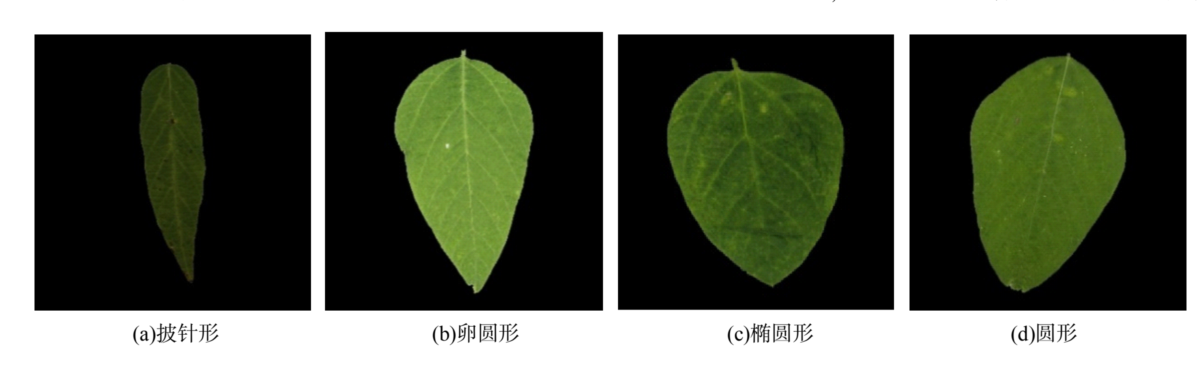
基于大豆叶片数据的特点，对于大豆叶片图像数据的分类识别，主要面临以下困难。

（1）数据图像易失真：目前，针对不同植物叶片的分类，大多是利用网上直接下载的数据集，虽然方便易用，但可能由于拍摄不能完全地垂直于叶片，导致大豆叶片形态产生非线性失真等问题。

（2）目标容易被遮挡：由于很多大豆叶片生长于同一位置，所以在田间采集的大豆叶片经常会被不同的叶片所遮挡。

（3）背景噪声影响大：直接将卷积神经网络用于不同植物叶片分类识别的处理方法虽然被广泛应用，但由于背景复杂，可能分类结果受背景噪声影响较大，只对单一数据集有效，不具备泛化能力。

针对这些困难，本文利用图像处理技术对采集的



(a)披针形 (b)卵圆形 (c)椭圆形 (d)圆形

图 1 大豆叶片的 4 种形态

Fig.1 Four forms of soybean leaves

大豆叶片进行感兴趣区域划分，叶片提取和标准化等处理，从而避免图像非线性失真，目标被遮挡和背景噪声等问题的影响，使训练的卷积神经网络具有更强的泛化能力。

2.1.2 图像采集

本文所使用的数据来自黑龙江省农业科学院大豆试验田中的大豆植株，在大豆盛花期，随机对来自不同大豆品种的植株中上部发育成熟的三出复叶进行采摘，以减少识别过程中不同品种之间大豆叶片形态的潜在差异，增加对不同种类大豆叶片形态识别的准确性。

将采摘好的 3 200 组大豆叶片依次放置在封闭的暗箱中进行图像采集。然后利用基于标记的分水岭分割方法对采集好的大豆叶片图像进行分割，每张分割好的大豆叶片图像包括三出复叶，利用基于标记的分水岭分割中创建的叶片掩模，使其只显示三出复叶中的顶小叶，将背景与其余叶片利用掩模覆盖变为黑色，最终得到 3 200 张单个大豆叶片图像。共分为 4 种类型：披针形，卵圆形，椭圆形和圆形。其中训练集与测试集按 7∶3 的比例划分，即 2 240 张作为训练图片，960 张作为测试图片。另外单独对训练集做数据增强，将训练集中的图像进行翻转和随机旋转，最终得到的训练集为 15 680 张大豆叶片图像。

2.2 实验方法

本文利用制作的暗箱采集装置对大豆叶片进行集中采集，再利用图像处理技术对大豆叶片图像进行感兴趣区域划分，叶片提取等处理，最后利用深度学习对大豆叶片进行分类识别。技术流程图如图 2 所示。

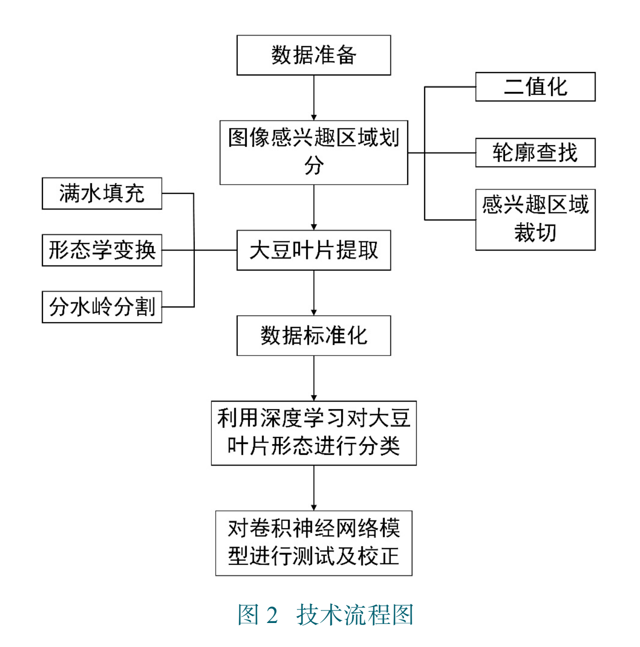


图 2 技术流程图

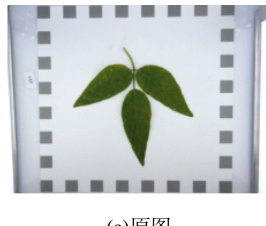
Fig.2 Technical flow chart

2.2.1 图像感兴趣区域划分

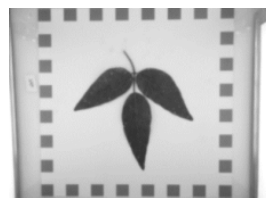
OpenCV 是由英特尔公司开发的，目前最常用的跨平台的计算机视觉库，可用于开发实时的图像处理、计算机视觉以及模式识别程序。因此本文利用 OpenCV 开源计算机视觉库对图像进行批量提取处理，对采集好的大豆叶片图像，进行灰度化处理以及高斯滤波处理，用以去除图像的噪声。然后进行二值化处理，得到二值图。最后利用轮廓检测，找到图像的最小外界四边形。根据最小外接四边形的 4 个顶点划分 ROI 感兴趣区域，去除暗箱中的背景对叶片分割的影响，得到提取好的叶片（图 3）。

2.2.2 叶片提取

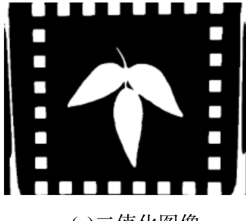
利用基于标记的分水岭分割算法对大豆叶片图像进行提取，基于标记的分水岭分割算法是在传统的分水岭分割算法的基础上，利用形态学变换确定要处理



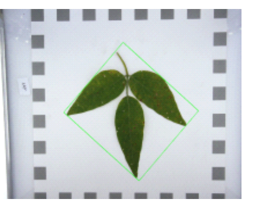
(a)原图



(b)灰度化经过高斯滤波



(c)二值化图像



(d)轮廓最小外接四边形



(e)裁切后图像

图 3 叶片图像提取过程

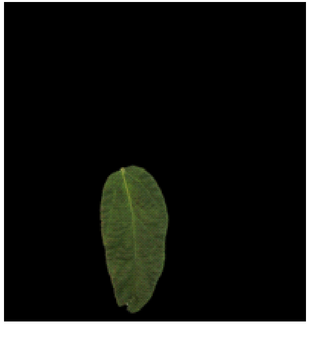
Fig.3 Leaf image extraction process

的图像的前景和背景区域。

首先将已经划分好感兴趣区域的大豆叶片图像看成拓扑平面，图像中每一点像素的灰度值表示该点的海拔高度，灰度值高的区域可以被看成是山峰，灰度值低的区域可以被看成是盆地。利用形态学变换中的开运算去除掉大豆叶片的根茎部分，通过腐蚀确定要处理的图像的前景区域，膨胀确定图像背景区域，进而对各个前景和背景区域依次进行标记，并在每个标记下灌入不同颜色的水，随着水位的升高，水通过流动将逐渐填充满所有空盆地，而来自不同的标记下的不同颜色的水将被水坝隔开。使其已经确定的各个前景和背景区域分别成为一个整体，从而根据被水坝隔离的各个盆地的不同标记，将完整的大豆叶片分成 3 个单个的大豆叶片，并将大豆叶片三出复叶中的顶小叶单独显示出来，其他叶片同背景一起使用掩模覆盖掉。从而避免因传统分水岭分割算法对图像灰度值的细微变化而产生的过度分割（图 4）。



(a)分割好的完整大豆叶片



(b)分割好的大豆三出复叶顶小叶

图 4　图像分割

Fig.4 Image segmentation

最后，利用 OpenCV 对分割好的大豆叶片图像进行轮廓检测，将检测到的轮廓边缘进行循环遍历，根据坐标筛选出轮廓边缘中最上，最下，最左，最右点的坐标，分别向四周增加 200 像素值，最后以上下长度和左右长度中的最大值作为标准化后的图像长度和宽度，裁剪成正方形，并统一改为 224×224 大小像素的图像，将其作为标准化后的图像输入深度学习进行训练，使其不受图像中前景位置的影响，提高训练的准确率和泛化能力。

2.3　特征提取与识别

本文采用 Densenet 网络模型进行大豆叶片的深层特征提取与形态识别，DenseNet 网络模型的特点是在保证网络中层与层之间最大程度的信息传输的前提下，直接将所有层连接起来，使网络的每一层都可以直接利用损失函数的梯度和最开始的输入信息，从而解决梯度消失现象，训练更深层的网络。简单讲，就是为了保证能够获得网络层之间的最大信息，将所有层进行互相连接，每一层的输入来自前面所有层的输出，如图 5 所示。尽可能地对特征进行极致利用，从而达到更少的参数，就可以训练更深、更准确、更有效的卷积网络的效果。

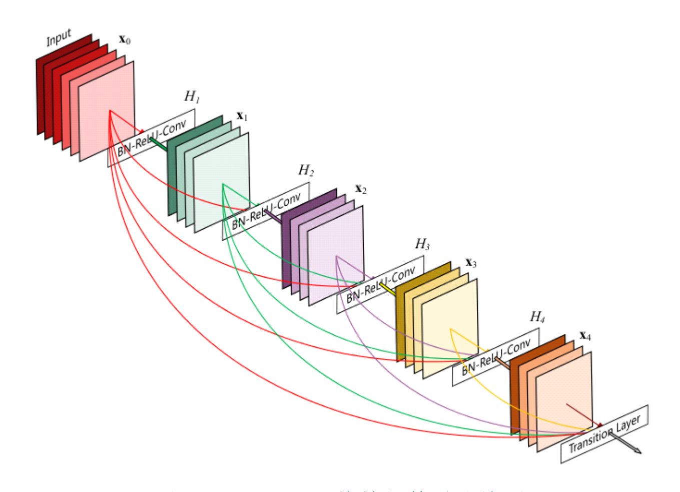


图 5　DenseNet 网络特征传递连接图

Fig.5 DenseNet network feature transfer connection diagram

2.3.1　网络模型结构

DenseNet 网络模型结构主要由 4 个 Dense Block 组成，每个 Dense Block 中，分别包含 6 个、12 个、64 个、48 个 Bottleneck 层，其中每个 Bottleneck 层都具有相同的层结构，即 BN-ReLU-Conv(1x1)-BN-ReLU-Conv(3x3)，它可以使各个 Dense Block 内的特征的尺寸统一，避免连接中出现尺寸的问题。并且为了使其网络更窄，参数更少，每个 Bottleneck 层最后的输出特征都为 48 个，这也保证了特征更有效的传递和网络更容易训练。

每个 Dense Block 之间的层，称为 Transition 层，主要完成卷积和池化的操作。其中，Transition 层由

BN 层、1x1 卷积层和 2x2 平均池化层组成，目的是，每个 Dense Block 结束后的输出信道个数很多，为了简化模型，需要用 1×1 的卷积核来降维，减小特征图的信道数量（图 6）。

2.3.2　网络训练流程

训练 DenseNet 对大豆叶片形态进行分类识别，首先要将处理好的 224×224 大小的不同形态大豆叶片输入到 DenseNet 网络中，网络初始的卷积层有 96 个卷积，卷积核大小为 7x7，步长为 2，填充为 3，然后进入 BN 层和卷积核大小为 3×3，步长为 2 的最大池化层。输出的特征将会依次进入 4 个 Dense Block 中，每个 Dense Block 之间的由 Transition 层相连接。经过 4 个 Dense Block 后，最后进入 BN 层，然后展开成 2 208 维的全连接层，利用 softmax 将其分为披针，卵圆，椭圆和圆 4 类（图 7）。

2.4　实验结果与分析

2.4.1　实验环境

本文使用的计算机为 Windows7 的 64 位操作系统，处理器为 Intel Core i5-8400，显卡为 GeForce GTX 1070Ti，内存为 8G。使用 Visual Studio 2015 作为开发环境和 OpenCV 开源计算机视觉库对采集的大豆叶片图像进行预处理。使用 Pytorch 框架，Jupyter Notebook 开发环境和 Python 语言用于在计算机上训练 DenseNet 网络模型以及测试预测结果。其中，DenseNet 采用 ReLU 激活函数，ReLU 激活函数具有收敛速度快和稀疏激活性的优点[5]。

损失函数采用了 Cross Entropy Loss 函数。即交叉熵损失函数，交叉熵越小，证明计算出的非真实分布越接近真实分布。该函数可以为 CNN 网络训练提供更

Layers	Output Size	DenseNet-121	DenseNet-169	DenseNet-201	DenseNet-264
Convolution	112 × 112	7 × 7 conv, stride 2			
Pooling	56 × 56	3 × 3 max pool, stride 2			
Dense Block (1)	56 × 56	[1 × 1 conv, 3 × 3 conv] × 6	[1 × 1 conv, 3 × 3 conv] × 6	[1 × 1 conv, 3 × 3 conv] × 6	[1 × 1 conv, 3 × 3 conv] × 6
Transition Layer (1)	56 × 56	1 × 1 conv			
	28 × 28	2 × 2 average pool, stride 2			
Dense Block (2)	28 × 28	[1 × 1 conv, 3 × 3 conv] × 12	[1 × 1 conv, 3 × 3 conv] × 12	[1 × 1 conv, 3 × 3 conv] × 12	[1 × 1 conv, 3 × 3 conv] × 12
Transition Layer (2)	28 × 28	1 × 1 conv			
	14 × 14	2 × 2 average pool, stride 2			
Dense Block (3)	14 × 14	[1 × 1 conv, 3 × 3 conv] × 24	[1 × 1 conv, 3 × 3 conv] × 32	[1 × 1 conv, 3 × 3 conv] × 48	[1 × 1 conv, 3 × 3 conv] × 64
Transition Layer (3)	14 × 14	1 × 1 conv			
	7 × 7	2 × 2 average pool, stride 2			
Dense Block (4)	7 × 7	[1 × 1 conv, 3 × 3 conv] × 16	[1 × 1 conv, 3 × 3 conv] × 32	[1 × 1 conv, 3 × 3 conv] × 32	[1 × 1 conv, 3 × 3 conv] × 48
Classification Layer	1 × 1	7 × 7 global average pool			
		1000D fully-connected, softmax			

图 6　DenseNet 网络模型参数图

Fig.6　DenseNet network model parameter diagram

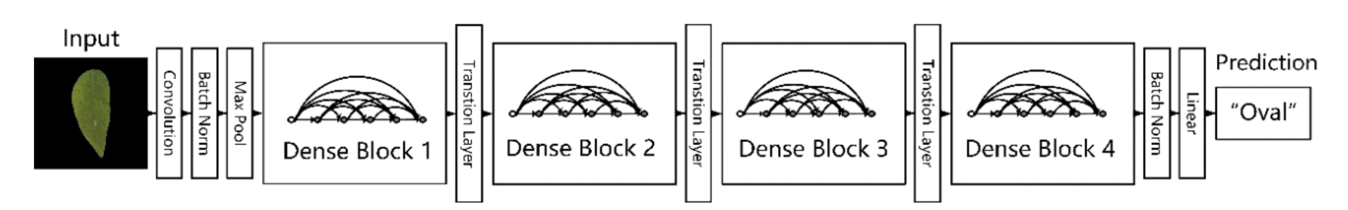


图 7　DenseNet 网络模型结构图

Fig.7　DenseNet network model structure diagram

为有效的惩罚目标函数，使得训练模型具有更好的泛化能力。

对 DenseNet 网络用 Momentum 优化函数进行训练。Momentum 优化函数是依据物理学的势能与动能之间能量转换原理提出来的，当 Momentum 动量越大时，其转换为势能的能量也就越大，就越有可能摆脱局部凹域的束缚，进入全局凹域。Momentum 动量主要用在权重更新的时候，可以让那些因学习率太大而来回摆动的参数，梯度能前后抵消，从而阻止发散。

一般，神经网络在更新权值时，采用如下公式：$w = w - \alpha \times dw$，引入 Momentum 后，采用如下公式：

$$v = mu \times v - \alpha \times dw$$

$$w = w + v$$

其中，v 初始化为 0，α 为学习率，mu 是设定的一个超变量，本文设定为 0.9。

2.4.2 评价标准

针对大豆叶片形态的分类识别，本文先选用传统机器学习方法中的最常用的支持向量机（SVM）和随机森林（RF）对大豆叶片形态进行识别，又在深度学习中选用经典的 AlexNet 网络和目前最为流行的 ResNet 网络对大豆叶片形态进行识别。最后并将不同方法的识别准确率与本文采用的 DenseNet 网络进行对比分析。每种大豆叶片形态的准确率计算如公式所示。

$$\theta = \frac{1}{M}\sum_{i=1}^{M} p_i$$

其中，式中 p_i 为 1 表示准确识别出图像形态种类，否则记为 0，M 表示测试数据的数量。

2.4.3 不同算法识别结果对比与分析

SVM 与 RF 作为传统的机器学习算法，需要利用图像处理技术对大豆叶片进行特征提取，所以本文提取大豆叶片图像的形态特征，纹理特征和颜色特征共 14 种特征作为 SVM 与 RF 的输入，其结果表明 SVM 与 RF 对大豆叶片形态的识别准确率低于本文使用的 DenseNet 网络。

而后本文又利用 AlexNet 网络和 ResNet 网络对经过图像处理的分割好的大豆叶片顶小叶图像作为卷积神经网络的输入，其结果表明 AlexNet 网络和 ResNet 网络对大豆叶片形态的识别准确率依然低于本文使用的 DenseNet 网络。

其中，本文使用的 DenseNet 网络识别精度结果如表 1 所示。表 1 结果表明，网络在 4 种不同大豆叶片的形态上，平均识别精度为 0.94。试验表明网络能够满足实际使用要求。

表 1　不同大豆叶片形状的识别精度

Table 1 Recognition accuracy of different soybean leaf shapes

大豆叶片类型	平均识别精度
披针形	0.98
卵圆形	0.96
椭圆形	0.88
圆形	0.93

不同算法对大豆叶片形态的识别准确率如表 2 所示，其结果表明 DenseNet 网络模型对于大豆叶片形态的分类识别相对于其他方法具有较好的效果。

表 2　不同机器学习方法对大豆叶片形状的识别精度比较

Table 2 Comparison of the recognition accuracy of different machine learning methods for soybean leaf shape

方法名称	识别准确率				
	披针	卵圆	椭圆	圆	平均值
SVM	0.95	0.96	0.78	0.48	0.80
AlexNet	0.99	0.82	0.68	0.90	0.85
ResNet	0.96	0.86	0.83	0.88	0.88
DenseNet	0.98	0.96	0.88	0.93	0.94

传统的机器学习算法由于需要对手动对大豆叶片的形态，纹理，颜色等特征进行提取，工作量大且特征提取困难。并且由于人为选择特征进行提取，受主观因素和自然环境影响较大，所以 SVM 与 RF 对大豆叶片形态的识别准确率较低。

而 AlexNet 网络和 ResNet 网络虽然直接将图像作为输入，不受主观因素和自然环境影响，但相对于 DenseNet 网络模型，AlexNet 网络由于使用 Dropout 正则化随机忽略一部分神经元，虽然避免了模型的过拟合，却也导致了一部分特征的丢失，所以对于比较易于分类的披针形有较好的识别能力，而对于其他 3 种形态则表现得不够理想。ResNet 利用在前面的特征层

与后面的特征层之间创建短连接。解决了深度学习网络越深，梯度消失现象越明显的问题，而 DenseNet 相对于 ResNet 选择直接将所有层连接起来，并通过在 Dense Block 的 3×3 卷积前面加入一个 1×1 的卷积，从而减少了输入的特征数量，既能降维减少计算量，又能融合各个通道的特征，使得网络更窄，参数更少，特征和梯度的传递更加有效，网络更加容易训练。

3 结　语

本文以黑龙江省农业科学院大豆试验田中的大豆植株豆叶片形态为例，对植物的叶片识别，疾病监测有可操作性，具有一定的推广作用，系统地阐述了利用深度学习对农业数据进行分类识别的全过程，通过分析大豆叶片形态图像特点后，基于深度学习开展了对大豆叶片形态的分类识别研究，我们已经看到，深度学习技术已经取代了使用手工特征训练的浅层分类器。只要有足够的数据可用于培训，深度学习技术就能高度准确地识别大豆叶片。采用了 DenseNet 模型，它适合用于普通网络的模型，该模型的优点是具有最佳的性能和最少的存储要求，但缺点是培训时间比较长，达到了较高的识别准确率 94%，并与其他图像识别方法进行对比分析，成功解决了传统方法在处理大豆叶片图像分类问题上耗时时间长，效率低下，且识别准确率不高等问题。可以满足农业图像数据分类的实际需要。未来的研究工作应努力收集范围广泛的大型且多样化的数据集发行以促进大豆叶片的识别研究，应侧重于开发可靠的背景去除技术以及合并其他形式的数据，提高大豆识别系统的准确性和可靠性。

参考文献：

[1] 姜恩波, 李娜. 中国开放政府农业数据分析与评价[J]. 农业图书情报学报, 2020, 32(10): 4–15.

JIANG E B, LI N. Analysis and evaluation of China's open government agricultural data[J]. Journal of agricultural library and information, 2020, 32(10): 4–15.

[2] 翁杨, 曾睿, 吴陈铭, 等. 基于深度学习的农业植物表型研究综述[J]. 中国科学: 生命科学, 2019, 49(6): 698–716.

WENG Y, ZENG R, WU C M, et al. A review of research on agricultural plant phenotypes based on deep learning[J]. Science China: Life sciences, 2019, 49(6): 698–716.

[3] 岑海燕, 朱月明, 孙大伟, 等. 深度学习在植物表型研究中的应用现状与展望[J]. 农业工程学报, 2020, 36(9): 1–16.

CEN H Y, ZHU Y M, SUN D W, et al. Application status and prospects of deep learning in plant phenotype research [J]. Transactions of the Chinese society of agricultural engineering, 2020, 36(9): 1–16.

[4] 袁培森, 黎薇, 任守纲, 等. 基于卷积神经网络的菊花花型和品种识别[J]. 农业工程学报, 2018, 34(5): 152–158.

YUAN P S, LI W, REN S G, et al. Chrysanthemum flower type and variety recognition based on convolutional neural network[J]. Transactions of the Chinese society of agricultural engineering, 2018, 34 (5): 152–158.

[5] WU J T, YANG G, YANG H, et al. Extracting apple tree crown information from remote imagery using deep learning[J]. Computers and electronics in agriculture, 2020, 174: 1–14.

[6] RAUF H T, SALEEM B A, LALI M, et al. A citrus fruits and leaves dataset for detection and classification of citrus diseases through Q 5 machine learning[J]. Data in brief, 2019, 26: 1–7.

[7] MYR J, KESKI-SAARI S, KIVINEN S, et al. Tree species classification from airborne hyperspectral and LiDAR data using 3D convolutional neural networks[J]. Remote sensing of environment, 2021, 256: 112322.

[8] DYRMANN M, KARSTOFT H, MIDTIBY H S. Plant species classification using deep convolutional neural network[J]. Biosystems engineering, 2016, 151: 72–80.

[9] 吕盛坪, 李灯辉, 冼荣亨. 深度学习在我国农业中的应用研究现状[J]. 计算机工程与应用, 2019, 55(20): 24–33, 51.

LV S P, LI D H, XIAN R H. The application research status of deep learning in agriculture in my country[J]. Computer engineering and applications, 2019, 55(20): 24–33, 51.

[10] Manavalan R. Automatic identification of diseases in grains crops through computational approaches: A review [J]. Computers and electronics in agriculture, 2020, 178: 1–24.

[11] 金瑛, 叶飒, 李洪磊. 基于 ResNet-50 深度卷积网络的果树病害智能诊断模型研究[J]. 农业图书情报学报, 2021, 33(4): 58-67.
JIN Y, YE S, LI H L. The intelligent diagnosis model of fruit tree disease based on resNet-50[J]. Journal of library and information science in agriculture, 2021, 33(4): 58-67.

Recognition and Classification of Deep Learning in Soybean Leaf Image Data Management

LU Lina[1,2], YU Xiao[1,2*]

(1. Business School, Shandong University of Technology, Zibo 255000;

2. School of Computer Science and Technology, Shandong University of Technology, Zibo 255049)

Abstract: [Purpose/Significance] We used to process soybean leaf data by looking at them and process data manually, but this method is very inefficient. In order to improve the classification accuracy and efficiency of soybean leaf images, further for storage and management of these images, we used the deep learning technique to make an in-depth study of text data and image data of soybean leaves for the image recognition and classification. The application of deep learning in agricultural data management mainly focuses on the image recognition and classification of plants and plant phenotypes in large-scale data, detection and classification of agricultural diseases and pests, detection and classification of crops and weeds, and prediction of crop yield. Through case analysis, our sample data demonstrated the application process of deep learning technology. [Method/Process] This paper systematically described the whole process of classification and recognition of agricultural data by using the deep learning technique. Through recognition and disease monitoring of plant leaves, the leaf morphology of soybean plants in the soybean experimental field of Heilongjiang Academy of Agricultural Sciences was taken as an example. We analyzed the image features of soybean leaf morphology, and carried out the classification and recognition research of soybean leaf morphology based on deep learning. Deep learning techniques have replaced shallow classifiers that use manual feature training and can identify soybean leaves with a high degree of accuracy as long as sufficient data are available for training. We adopted DenseNet model, which is suitable for common network model. The advantages of this model are that it has the best performance and the least storage requirements. First， we selected support vector machine (SVM) and random forest (RF) in traditional machine learning methods to identify soybean leaf morphology. Second, AlexNet and ResNet were selected to identify soybean leaf morphology. Finally, the recognition accuracy of different methods were compared with the DenseNet network adopted in this paper. [Results/Conclusions] Through the training of DenseNet model, the recognition accuracy of 94% was achieved, which successfully solved the problems of long time, low efficiency and low recognition accuracy of traditional methods in processing image classification of soybean leaves, and could meet the actual needs of agricultural image data classification. Future research efforts will strive to collect a wide range of large and diverse data sets to facilitate soybean leaf recognition, and focus on developing reliable background removal techniques and incorporating other forms of data to improve the accuracy and reliability of soybean leaf recognition systems.

Keywords: deep learning; agricultural science data; data classification; image recognition